Statistical Models for Spot Air Quality Forecasts (O3 and PM10) in British Columbia

Jon Y. Wang, Edward Lord, Alex Cannon, Gregg Walters
Meteorological Service of Canada
Pacific and Yukon Region, Environment Canada
201 - 401 Burrard Street, Vancouver, BC V6C 3S5

Keywords: Ozone, Particulate matter, Air quality forecast

1. Abstract

Multiple linear regression models were developed for spot forecasts of 1-hour ozone (O3) and 24-hour particulate matter with diameter less than 10um (PM10) at 28 sites in the interior British Columbia, Lower Fraser Valley and Vancouver Island of British Columbia. The models were trained based on datasets from 2000 to 2004 at each site including meteorological variables from Scribe and XScribe matrices from Canadian Meteorological Centre, air quality variables (antecedent O3 and PM10) from the Greater Vancouver Regional District air quality monitoring network, and seasonal variable (sine function of Julian day). Seventeen models over the next 48 hours (0-h, 3-h, 6-h, …, 48-h) for O3 and PM10 at each of the 28 sites were implemented for air quality forecasts (O3 and PM10) covering 16 regions in British Columbia. Cross-validation showed that the average *bias*, *MAE*, and *RMSE* were -0.018(ppb), 5.590(ppb), and 7.246(ppb) for O3 model, and -0.001(ug/m^3), 2.988(ug/m^3), and 4.051(ug/m^3) for PM10 model. The models are currently used operationally in the Pacific and Yukon Region of Environment Canada. The official forecast (FLCN40 CWVR) is issued to the public daily by the Pacific Storm Prediction Centre located in Vancouver, British Columbia.

2. Introduction

The Pacific Weather Centre began forecasting ground level ozone for Greater Vancouver and the Fraser valley sites in 1992. The original air quality bulletin was only issued from May through September reflecting the time of year when ground level ozone concentrations in British Columbia (BC) can become a concern. In 2003 the forecast was expanded to include 24 hr average PM10 concentrations and 2 forecast sites in the BC Interior, Kamloops and Kelowna. On March 1, 2004 the air quality program was expanded to run all year long; as well as adding forecasts for Nanaimo, Victoria and Prince George. At present two forecasts (FLCN40 CWVR AQI Forecast and FLCN39 CWLW Smoke Control Forecast) relating to air quality are issued at the Pacific Storm Prediction Centre (PSPC). The main forecast is the FLCN40 CWVR issued at 6:00AM daily which is based on our statistical models. This bulletin forecasts an AQI value and category for 16 regions including 28 sites in BC. Of those 16 regions, 13 are in the Inner South Coast and 3 are located in the BC Interior (Table 1and Figure 1).

Table 1 Regions and sites covered by the models

Region	Site
VANCOUVER	DOWNTOWN VANCOUVER; KITSILANO
NORTH SHORE	SECOND NARROWS; MAHON PARK
RICHMOND DELTA	RICHMOND SOUTH; VANCOUVER AIRPORT
BURNABY NEW WES	KENSINGTON PARK; BURNABY MOUNTAIN; BURNABY SOUTH
NORTHEAST	PORT MOODY; COQUITLAM
SURREY N DELTA	NORTH DELTA; SURREY EAST
RIDGE MEADOWS	PITT MEADOWS; MAPLE RIDGE
LANGLEY	LANGLEY
ABBOTSFORD	ABBOTSFORD
CHILLIWACK	CHILLIWACK
HOPE	HOPE AIRPORT
VICTORIA	SAANICH STELLYS CR; VICTORIA ROYAL ROA; VICTORIA TOPAZ
NANAIMO	NANAIMO LABIEUX ROA
KELOWNA	KELOWNA COLLEGE
KAMLOOPS	KAMLOOPS BROCKLEHU
PRINCE GEORGE	PRG PLAZA 400; PRG BC RAIL WAREHO; PRG GLADSTONE SCH

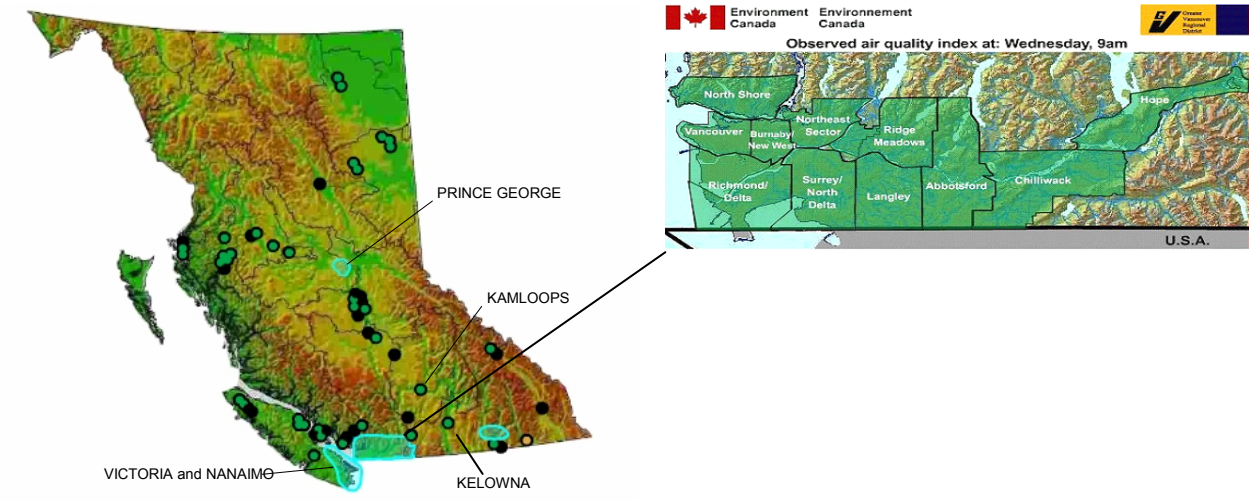

Figure 1 Locations of the regions and sites covered by the models

3. Data and Method

Dataset of 2000 to 2004 at each site includes meteorological variables (Table 2) from Scribe and XScribe matrices from Canadian Meteorological Centre (CMC), air quality variables (antecedent O3 and PM10) from the Greater Vancouver Regional District (GVRD) air quality monitoring network, and seasonal variable (sine function of Julian day). Seasonal Development was defined as:

$$SD = \sin\left[\left(Julian\ Day - Julian\ Day\ March\ 21\right)2\pi / 365\right]$$

Each dataset was extended using the following procedure:

- Each column (variable) in the dataset was extended to three new columns using functions of square and square root, ending a new dataset: V_1, V_1^2, $V_1^{0.5}$, V_2, V_2^2, $V_2^{0.5}$ … V_n, V_n^2, $V_n^{0.5}$
- The new dataset was further extended using a combination of two variables, ending the final dataset for training the models: V_1, V_1^2, $V_1^{0.5}$, V_2, V_2^2, $V_2^{0.5}$ … V_n, V_n^2, $V_n^{0.5}$, V_1xV_1, $V_1xV_1^2$, $V_1xV_1^{0.5}$, V_1xV^2, $V_1xV_2^2$, $V_1xV_2^{0.5}$ … $V_n^{0.5}xV_n$, $V_n^{0.5}xV_n^2$, $V_n^{0.5}xV_n^{0.5}$
- 17 models at each site were trained based on the final dataset using stepwise regression approach

Outputs from the models are grouped into "GOOD", "FAIR" and "POOR" air quality categories. Good air quality is defined as a maximum ozone reading of 51 ppb or less, fair air quality of 52 to 82 ppb, and poor air quality of greater than 82 ppb. These three categories are defined in reference to the Canadian desirable air quality objective for ozone of 52 ppb and the Canadian maximum acceptable air quality objective for ozone of 82 ppb (http://www.atl.ec.gc.ca/airquality/whatis_index_e.html). Air quality can also be determined according to PM10 values (Table 3).

The bias *Bias*, mean absolute error *MAE*, and root mean square error *RMSE* were used to evaluate the performance of each of the models. These measures are defined as follows:

$$Bias = \frac{\sum_{i=1}^{N}(P_i - O_i)}{N}$$

$$MAE = \frac{\sum_{i=1}^{N} abs(P_i - O_i)}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}}$$

where N equals to the number of observations, P_i and O_i are the ith predicted value and observation, and $\overline{P}$ and $\overline{O}$ are the average of the model predictions and observations.

Table 2 Variables used for training the models

Variables from Scribe matrices	Variables from XScribe matrices
UTC Cumulative forecast valid time in hours	Mean Sea Level Pressure (0.1 hPa)
UTC forecast valid time in hours	Lifted index (nearest unitity)
Climatological max/min	PBL Average Wind Direction (degrees)
Spot time temperature forecast	PBL Average Wind Speed (km/h)
Climatological frequency of pcpn amount of 0.2 mm or more in 6 hr	Instantaneous VIS SFC incident flux (0.1 W/m2)
Climatological frequency of pcpn amount of 0.2 mm or more in 12 hr	Surface Pressure (0.1 hPa)
Climatological frequency of pcpn amount of 10.0 mm or more in 12 hr	Model Temperature at 925 mb (oC)
Cloud coverage (in tenth)	Model Dew Point at 925 mb (oC)
Pop over 6 hr period	Model Temperature at 850 mb (oC)
Pop over 12 hr period	Model Dew Point at 850 mb (oC)
Probability of getting 10.0 mm or more over 12 hr period	Model 925 mb Wind Direction (degrees)
Precipitation amount (in tenth of mm)	Model 925 mb Wind Speed (km/h)
Showalter index	Model 850 mb Wind Direction (degrees)
Vertical velocity at 850 mb in µbar/sec	Model 850 mb Wind Speed (km/h)
Conditional precipitation type	
Surface dew point depression	**Other variables**
1000-500 mb thickness	Antecedent O3
850-700 mb thickness	Antecedent PM10
1000-850 mb thickness	Julian Day
Wind direction in degrees	Weekday
Wind speed in km/h	
Total cloud opacity (in tenth)	
Diagnostic probability of precipitation over a 6 hours period	
Diagnostic probability of precipitation over a 12 hours period	
Surface temperatures (in oC)	
Temperatures at the level nearest to 0.97 (in oC)	
Dew point depression at the level nearest to 0.97 (in tenth oC)	
Wind speed at the level nearest to 0.932, roughly 600 meters (in km/hr)	
Convective precipitation quantity amount (in tenth of mm)	
Snow on ground (in cm.)	

Detailed descriptions of Scribe/XScribe variables can be found at http://iweb.cmc.ec.gc.ca/cmdw/wx_element/scribe/Guide_Scribe/Matrice/matriceE.html.

Table 3 Air quality index (AQI)

O3(ppb)	PM10(ug/m3)	AQI	Category
0-51	0-25	0-25	Good
52-82	26-50	26-50	Fair
83-153	51-100	51-100	Poor
>153	>100	>100	Very Poor

The *Bias* measure describes the degree to which the model over-predicts or under-predicts the dependent variable (hourly maximum ozone concentrations). The *MAE* is the sum of absolute errors divided by the number of observations, where absolute error is the absolute value of the difference between predicted value and observation. The closer the *MAE* is to zero, the more accurate the forecasts are. The *RMSE* measure describes the average error ignoring the sign of each error. As was discussed by Lord (2002), *RMSE* is sensitive to large errors due to the squaring of each error; therefore, a large *RMSE* indicates the occurrence of large errors. An *RMSE* equal to zero means a perfect prediction.

4. Model Implementation and Usage

To use the models, four input files (Scribe, XScribe, O3, and PM10) are necessary; these files are updated daily. In particular, the Scribe and XScribe matrices have fixed format and the models retrieve necessary data as inputs automatically; while on the other hand, the pollutant files can be modified by forecasters before input into the system (Figures 2 and 3). Input variables vary from site to site depending on the trained model at specific location.

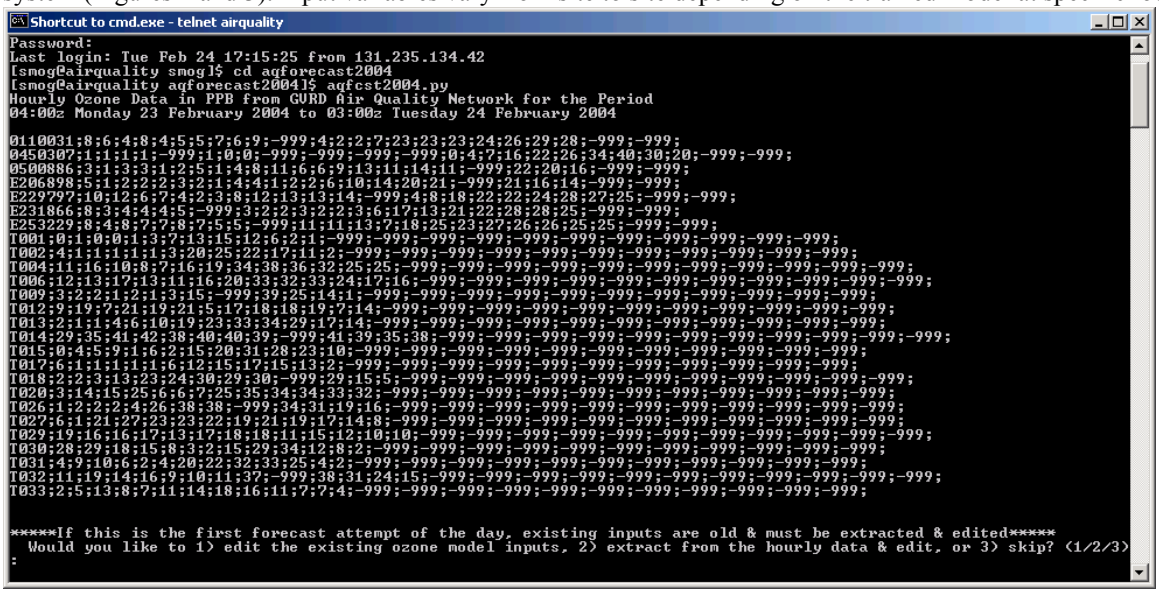

Figure 2 O3 hourlies retrieved from the most resent observations from GVRD (PM10 hourlies will be shown in the same format each time when the model is run.)

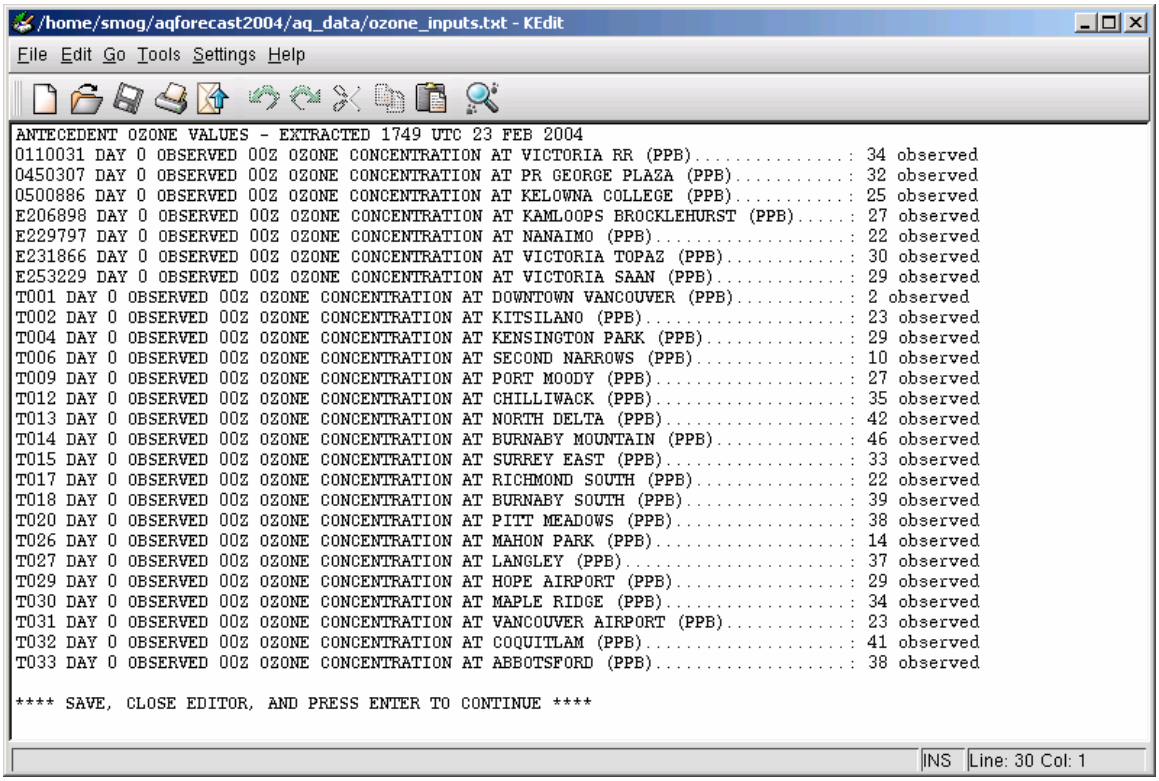

Figure 3 Editor window for modifying or changing O3 values (Similar window will be shown for PM10 values each time when the model is run.)

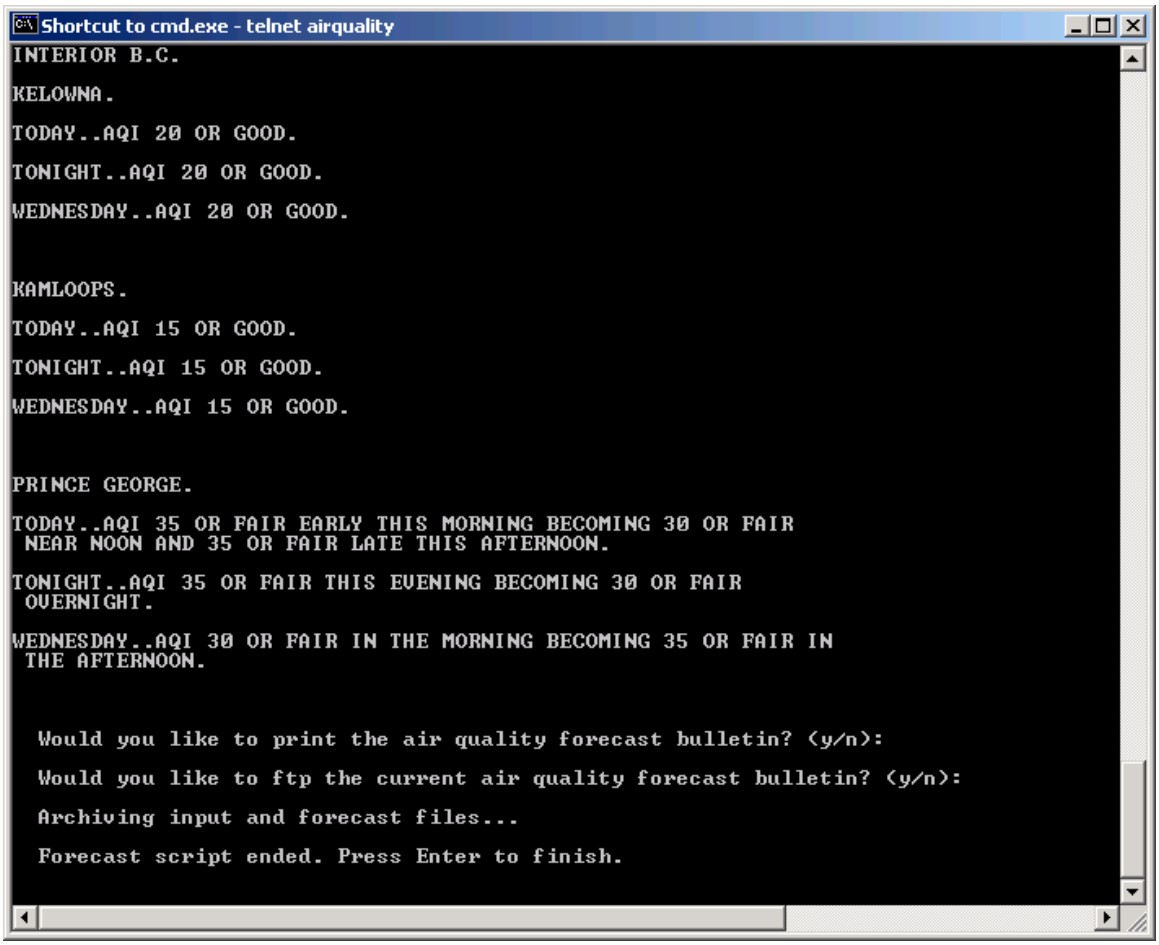

Figure 4 Output of some of the text forecasts

The models generate 1-hour average O3 forecasts (ppb) and 24-hour average PM10 forecasts (ug/m3) for different sites. Site specific AQI forecasts are computed by comparing the O3 and PM10 values at the site. The system then extracts the today, tonight, and tomorrow air quality index (AQI) spot forecasts from a give region's 17 forecasts over 48 hours (0-h, 3-h, 6-h, …, 48-h), and gives text forecasts for each region (Figure 4).

5. Results and Conclusions

AQI cross-validation results based on 2001-2004 data are shown in Figures 5. The statistical models outperform AQI persistence and 24-hour lag (from antecedent O3 and PM10). The models were implemented and run in March 2004. Model performance during the last year is shown in Tables 4 and 5.

AQI forecasts from the models are reasonably accurate and better than those from persistence, 24-hour lag, and CHRONOS -- The Canadian Hemispheric and Regional Ozone and NOx System (forecasts from CHRONOS not shown here).

On average, the *bias*, *MAE*, and *RMSE* for forecasting O3 were -1.034 (ppb), 7.107 (ppb), and 9.303 (ppb), respectively. Average *bias*, *MAE*, and *RMSE* for forecasting PM10 were 0.556 (ug/m3), 3.976 (ug/m3), and 5.128 (ug/m3), respectively.

Large errors may be found at some sites (O3 at SANNICH STELLYS CR and PM10 at PRG BC RAIL WAREHO). The main reason is probably due to the shortage of the data record used for training the models at these sites.

Our future work will involve updating the models using information available from most recent data. Environment Canada is planning on implementing a new health based air quality program in 2007 to replace its current air quality forecast program based on the maximum concentration of either PM10 or O3. The new air quality health index forecast will be based on concentrations of PM2.5, NO2, SO2 and ground level ozone. A pilot project is being developed that will be implemented later this year for Kamloops, Vernon and Kelowna in the BC Interior.

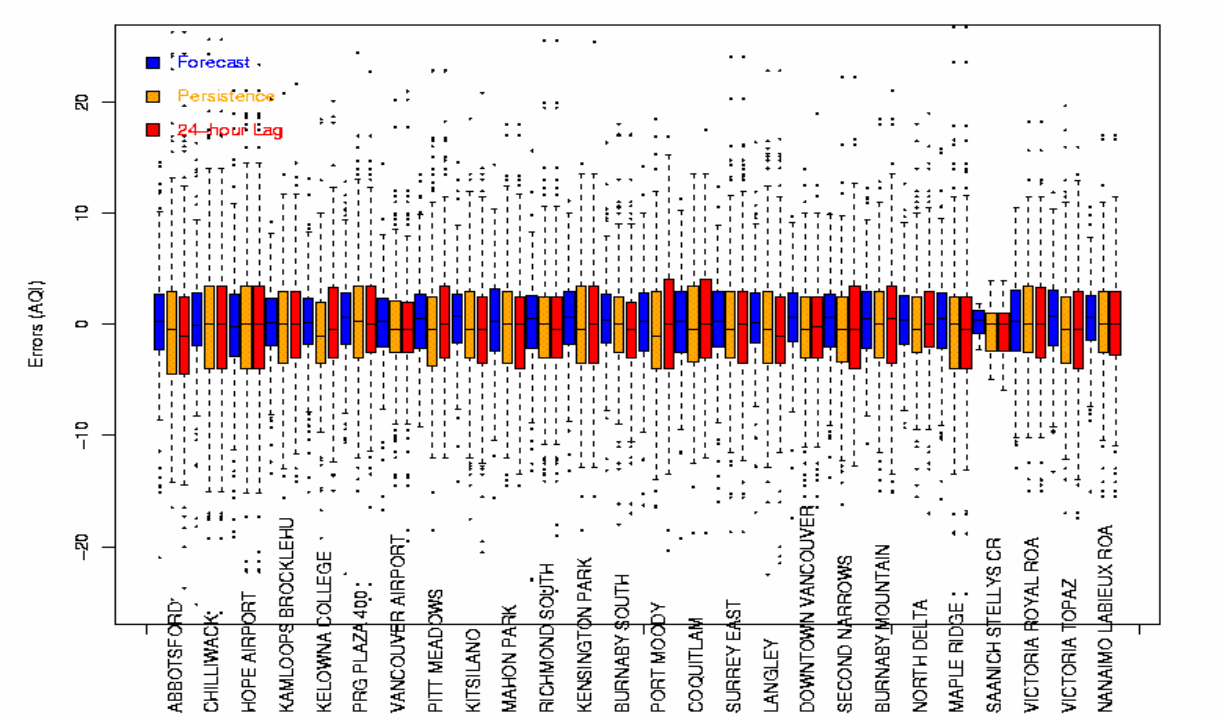

Figure 5 AQI error comparison between the model, persistence, and 24-hour lag

Table 4 Evaluation on PM10 forecast in 2004

Model evaluation based on model performance in 2004 (PM10 in ug/m^3)						
	Winter Model			**Summer Model**		
Site	bias	MAE	RMSE	bias	MAE	RMSE
KITSILANO	0.285	3.498	4.902	0.101	2.987	3.762
MAHON PARK	1.486	2.733	3.508	0.762	2.900	3.702
RICHMOND SOUTH	-0.010	3.252	4.173	0.688	2.880	3.578
VANCOUVER AIRPORT	0.803	3.120	3.798	-0.099	2.777	3.520
KENSINGTON PARK	1.400	2.782	3.560	-0.009	2.957	3.916
BURNABY SOUTH	1.413	2.905	3.621	0.904	3.099	3.96
PORT MOODY	1.242	2.795	3.779	0.505	3.328	4.356
SURREY EAST	-0.570	3.349	4.387	1.015	2.963	3.727
PITT MEADOWS	0.785	3.292	4.100	0.595	2.895	3.736
LANGLEY	2.550	3.813	4.779	0.491	2.573	3.385
ABBOTSFORD	1.725	3.651	4.727	0.929	3.458	5.368
CHILLIWACK	0.646	3.680	4.624	0.983	3.100	4.122
HOPE AIRPORT	2.754	3.560	4.375	0.293	3.056	4.072
KAMLOOPS BROCKLEHU	1.860	5.292	7.017	0.190	3.317	5.013
KELOWNA COLLEGE	-2.073	5.000	6.423	-0.282	3.134	4.847
PRG PLAZA 400	1.932	8.306	10.41	0.229	4.916	6.351
PRG BC RAIL WAREHO	-3.098	15.75	20.36	-1.438	8.681	12.20
PRG GLADSTONE SCHO	0.595	4.097	5.481	0.427	3.246	4.190

Table 5 Evaluation on O3 forecast in 2004

Model evaluation based on model performance in 2004 (O3 in ppb)						
	Winter Model			Summer Model		
Site	bias	MAE	RMSE	bias	MAE	RMSE
DOWNTOWN VANCOUVER	0.051	4.508	6.353	-1.748	5.132	7.132
KITSILANO	0.627	6.463	9.020	-1.200	6.365	8.390
SECOND NARROWS	2.268	6.664	9.287	-0.885	5.925	7.904
MAHON PARK	0.407	7.387	9.362	-1.973	7.341	9.669
RICHMOND SOUTH	-1.334	6.442	9.061	-1.204	6.473	8.488
VANCOUVER AIRPORT	-0.662	7.132	9.477	-1.259	6.916	9.453
KENSINGTON PARK	-2.046	5.932	7.875	-1.376	6.713	8.641
BURNABY MOUNTAIN	-7.453	8.725	10.95	-3.014	6.899	8.988
BURNABY SOUTH	-1.493	5.391	6.982	-0.714	6.293	8.162
PORT MOODY	-0.956	5.727	7.955	-1.692	6.626	9.220
COQUITLAM	0.795	6.024	8.016	-2.311	7.096	9.292
NORTH DELTA	-0.996	6.296	8.093	-2.812	6.904	8.974
SURREY EAST	-0.277	6.116	7.941	-1.571	6.276	8.096
PITT MEADOWS	0.961	6.059	7.740	-0.522	6.756	8.986
MAPLE RIDGE	-0.791	5.303	7.171	-0.751	6.169	7.988
LANGLEY	0.277	6.668	8.871	0.949	6.742	10.95
ABBOTSFORD	0.937	6.740	8.600	-0.897	6.950	9.675
CHILLIWACK	-0.260	6.762	8.949	-0.080	7.050	9.180
HOPE AIRPORT	0.919	7.828	9.840	0.835	7.622	9.603
SAANICH STELLYS CR	-13.49	13.86	17.23	-12.04	12.50	15.40
VICTORIA ROYAL ROA	-2.498	7.482	9.907	-3.996	8.027	10.27
VICTORIA TOPAZ	-0.814	6.957	9.302	-2.174	7.063	9.218
NANAIMO LABIEUX RO	5.801	10.09	11.91	10.47	11.83	13.60
KAMLOOPS BROCKLEHU	-0.041	6.673	9.266	1.893	6.874	8.97
KELOWNA COLLEGE	0.086	6.568	8.645	-2.180	7.557	9.596
PRG PLAZA 400	-3.091	8.885	11.25	-0.414	6.782	8.861

6. References

Lord, E. R., 2002, An Evaluation of Objective Models Used to Forecast Summertime Ground-Level Ozone in the Lower Fraser Valley of British Columbia, Internal eReport #2002-003, Pacific Weather Centre, Environment Canada